

With good marks -

*been in
STAT
11/2/55*

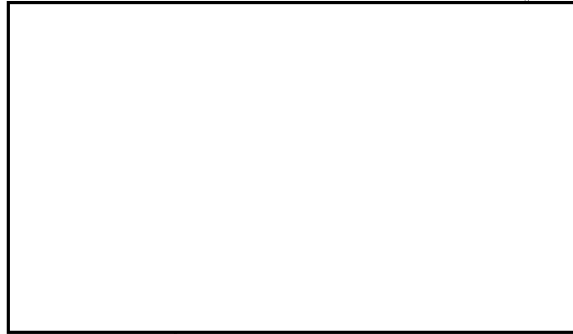
Reprinted from SCIENCE, October 21, 1955, Vol. 122, No. 3173, pages 745-748.

*60-00050A
folder
#4*

Mechanical Translation

New Challenge to Communication

Jacob Ornstein



STAT

Dissem 3-1

Reprinted from SCIENCE, October 21, 1955, Vol. 122, No. 3173, pages 745-748.

Mechanical Translation

New Challenge to Communication

Jacob Ornstein

One of the thorniest problems in present-day communication is the translation of important writings, especially in the fields of science and technology, from one language to another. The machine age has done little to alter the translation process, which today, as it was centuries ago, must be laboriously performed at considerable expense by human beings. Moreover, growing concern over the dangerous lag between the appearance of important works in other languages and their translation into English and other tongues has within the past decade focused serious attention on the problem of mechanical translation. Until recently, however, mechanical translation has remained one of those human dreams the realization of which was relegated to some point in the unpredictable future.

It was with the development of the electronic computer that real hope came to be felt concerning the possibility of mechanical translation. The man who first envisioned the possibility of translation by electronic means was apparently Warren Weaver, director of the Natural Sciences Division of the Rockefeller Foundation. First of all in 1945, and more concretely in a memorandum dated 15 July 1949, Weaver raised the question of the feasibility of designing a computer-like machine capable of translating from one language to another. This memorandum was circulated among a number of

linguists and scientists; their reactions ranged from high optimism to complete skepticism. Weaver's concepts, nevertheless, aroused considerable interest and stimulated preliminary research in the field.

The problem of mechanical translation is being approached by a growing number of scholars, although much of the research is in the realm of speculation and theory. Among mathematicians, one may mention such figures as Yehoshua Bar-Hillel, formerly of Massachusetts Institute of Technology and now of Hebrew University, whose articles on mechanical translation were accorded considerable space in the 5 April 1954 issue of *Time*, the weekly news magazine. Important research is being carried on at Massachusetts Institute of Technology by Victor H. Yngve of the Research Laboratory of Electronics and William N. Locke, professor of modern languages. Anthony Oettinger of Harvard University's Computation Laboratory has also been conducting experiments. Erwin Reifler of the University of Washington has concerned himself with the elaboration of a mechanical dictionary. An active group, headed by William Bull and Victor Oswald, is working on problems of mechanical translation at the University of California at Los Angeles.

The first successful experiment in performing mechanical translation was the

result of a joint project undertaken by Georgetown University's Institute of Languages and Linguistics, Washington, D.C., and the International Business Machines Corporation. Considerable publicity followed their demonstration at I.B.M. headquarters in New York on 7 January 1954 of the translation of more than 60 sentences from Russian into English. An invited group of government officials, linguists, and scientists watched a typist who knew no Russian type the sentences, which had been transliterated into the Roman alphabet, on an electric card punch and feed them into the "electronic translator," which produced accurate English translations. The sentences were from the workaday fields of science, technology, communications, and international affairs. The following are a few examples of the transliterated Russian sentences and the English equivalents:

Myezhdunarodnoye ponyimaniye yavlyayetsya vazhnim faktorom v ryesheniyi polyityicheskykh voprosov.

International understanding constitutes an important factor in decision of political questions.

Dorogi stroyatsya yiz byetona.

Roads are constructed from concrete.

Komandir poluchayet svyedyeniya po tyelyegrafu.

A commander gets information over a telegraph.

Vyelyichina ugla opredelyayetsya otnosheniyem dliny dugi k radiusu.

Magnitude of angle is determined by the relation of length of arc to radius.

Obrabotka povyshayet kachestvo nyefti.

Processing improves the quality of crude oil.

It is revealing to consider in some detail the background of the Georgetown-

The author is on the staff of the U.S. Department of Agriculture Graduate School.

Each instruction written in ordinary English had to be converted into terms of detailed programs for the individual computations of the machine. A single sentence of verbal instructions might require 100 computations on the part of the machine. The program steps for the machine, prepared by Sheridan, totaled about 2400. The dictionary syntax flow chart shown in Fig. 1 gives an idea of what the programming represents.

The designers also elaborated three codes that indicated to the machine which of the six rules of operational syntax applies to each word inserted for translation. These rules, or rule-tags, which were developed by Garvin, govern the transposition of words, choice of meanings where there are several alternatives, omission of unnecessary words, and insertion of necessary words. Their role, therefore, is to order the units so that the words will not come out a mere jumble.

Let us take a few examples of the functioning of the rules. The Russian words *gyeneral mayor* indicate a rank roughly equivalent to major general. Obviously the two words must be reversed. The switch is assured in advance by attaching the rule-tag 21 both to the Russian word *gyeneral* and to its translation in the bilingual glossary stored in the machine, and by attaching rule-tag 110 to the Russian word *major* and its translation. According to the stored instructions, whenever a rule-tag 110 is encountered in the glossary, it is necessary to go back and look for a rule-tag 21. If a 21 is found, the two translations must be printed in reverse order. Thus the translation *major of mayor* is printed before the translation *general of gyeneral*.

One more illustration is worth noting. The Russian word *o* can mean either *about* or *of*. In the Russian-English glossary *nauka* has affixed to it the rule-tag 242 and *o* carries the rule-tag 141. The instructions indicate to the machine that whenever rule-tag 141 is encountered, it is necessary to go back and search for 241 or 242. If 241 is found, the first English translation is selected and both words are printed in the order in which they appear in the Russian sentence. If 242 is encountered, the second English meaning is selected. Consequently, the computer reads the 141, looks for and finds 242, chooses the second meaning given for *o*, which is *of*, and prints correctly *science of*.

Table 1 contains the three codes and the six Garvin rules of syntax; Fig. 2 illustrates how a source sentence is translated, analyzed, and arranged in correct English word order by the converted I.B.M. 701 electronic data-processing machine.

The Russian language was chosen by the designers, but any one of the 2000-odd languages of the world might have

been chosen. One reason for the selection of Russian was its strategic importance; another was the fact that there is a relatively small number of persons competent to handle Russian, while the accumulation of untranslated works in that tongue continues to increase at an alarming rate. This accumulation constitutes an overflowing reservoir of data about the Soviet Union—books, newspapers, and journals available in Russian to any interested party. Moreover, the Georgetown linguists felt that in view of the highly inflected nature of the language, which is similar to Greek and Latin in its multiplicity of endings, a mechanical translator capable of handling Russian would yield significant experience for dealing with other complicated languages.

A preponderance of sentences from scientific and technical fields was chosen for the trial run, not only because of obvious timeliness, but also because writing in scientific and technical fields is done with words having highly specialized and precise meanings. Hence, if a word appears in a certain context, the chances of its having a single, unambiguous meaning are extremely high. The same possibilities for accurate prediction occur in other fields of technical writing such as medicine and engineering. Consequently, the Georgetown linguists assume that electronic translation will begin with separate dictionaries for each technical area; as experience with these areas grows, enough will be learned eventually to permit accurate translation of our common everyday language as well.

Two major problems exist for designers of mechanical translators. The first is that of reducing the number of synonyms of any given word for storage in the mechanical dictionary. This is rendered difficult by the fact that compilers of conventional dictionaries have persisted in giving an unnecessarily large number of equivalents for each entry. The experience with the electronic translator has indicated that improved selective techniques based on the principles of present-day linguistics will help provide answers to this problem.

The second difficulty is that of "operational syntax," which in mechanical translation parlance means the obtaining of an intelligible output sequence from an input with a different sequence of elements. This can be solved only by preparing detailed linguistic instructions that in turn must be converted into electronic instructions to the computer. Only by such means can machines produce intelligent translations relatively free of ambiguities.

The converted 701 computer with its 250-word vocabulary and six Garvin rules of operational syntax is still too limited to be effective for full-scale translation. This means that designers must

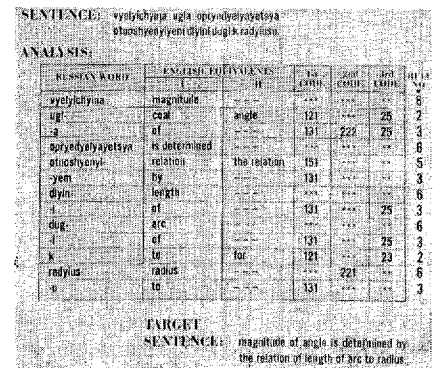


Fig. 2. Diagram illustrating how a source sentence is translated, analyzed, and arranged into the target sentence in correct English word order by the I.B.M. 701 electronic data-processing machine. See Table 1 for rules of operational syntax.

Table 1. Rules of operational syntax to accompany Fig. 2

Rule 1: Rearrangement. If first code is 110, is third code associated with preceding complete word equal to 21? If so, reverse order of appearance of words in output (that is, word carrying 21 should follow that carrying 110); otherwise, retain order. In both cases, English equivalent I associated with 110 is adopted.

Rule 2: Choice—Following Text. If first code is 121, is second code of the following complete, subdivided, or partial (root or ending) word equal to 221 or 222? If it is 221, adopt English equivalent I of word carrying 121; if it is 222, adopt English equivalent II. In both cases, retain order of appearance of output words.

Rule 3: Choice—Rearrangement. If first code is 131, is third code of preceding complete word or either portion (root or ending) of preceding subdivided word equal to 23? If so, adopt English equivalent II of word carrying 131 and retain order of appearance of words in output; if not, adopt English equivalent I and reverse order of appearance of words in output.

Rule 4: Choice—Previous Text. If first code is 141, is second code of preceding complete word or either portion (root or ending) of preceding subdivided word equal to 241 or 242? If it is 241, adopt English equivalent I of word carrying 141; if it is 242, adopt English equivalent II. In both cases, retain order of appearance of words in output.

Rule 5: Choice—Omission. If first code is 151, is third code of following complete word or either portion (root or ending) of following subdivided word equal to 25? If so, adopt English equivalent II of word carrying 151; if not, adopt English equivalent I. In both cases, retain order of appearance of words in output.

Rule 6: Subdivision. If first code associated with a Russian dictionary word is ***, then adopt English equivalent I of alternative English language equivalents, retaining order of appearance of output with respect to previous word.

embark upon a new phase of development, seeking both to amplify the vocabulary and to increase the ability of the machine to handle a wider range of operational syntax. The six rules will have to be vastly increased in number. Dostert estimates that about 100 rules would be needed to govern a vocabulary of 20,000 words. All this will require a great deal of additional linguistic research, as well as a special study, from the mechanical-translation point of view, of the source language vis-à-vis the target language. From the technical standpoint, this will involve increasing the machine storage and possibly the devising of special circuits. It will be necessary to analyze translation samples of increasing length and complexity and to adjust the results progressively after each analysis. The validity of the programming must continually be tested against machine equipment.

Looking realistically at the electronic translator, its developers recognize that the I.B.M. 701, which costs about \$500,000, is "overdesigned" for language translation; it has many functions not essential to this task that were built in to solve problems in astronomy and physics. The bulkiness of the 11 units, which occupy roughly the same area as a tennis court, is another drawback. According to Hurd, I.B.M. is considering the development of a machine exclusively intended

for translation. However, this cannot be accomplished until the Georgetown language specialists have elaborated additional instructions and specifications based on the experience that has been gained in the first phase of the joint project.

The time is still distant when it will be possible to insert a Russian scientific book into a translating machine and to receive from it an acceptable translation. It is highly doubtful that any machine could ever be devised that is capable of performing satisfactory translation of works of fiction or writings in which fine shades of meaning and subjective interpretations are involved. The translation of such a work as Milton's *Paradise Lost* or Tolstoy's *War and Peace* is in itself a creative act. Nor would mechanical devices ever eliminate the need for professional human translators. On the contrary, the latter would be freed from dull, routine hackwork and could devote themselves to the translation of works of literary and artistic merit that are more challenging in nature.

Neil MacDonald, describing the Georgetown-I.B.M. project in his article "Language translation by machine," which was published in the February 1954 issue of *Computers and Automation*, pointed out that the search for the solution of the translation problem brought to light many new facts that will

tend to bridge the gap between the humanities and science. For example, it was found that the formulation of the logic required to convert word meanings properly, even in a small segment of two languages, necessitated as many instructions to the computer as are required to simulate the flight of a guided missile. He predicted that in the future

"Linguists will be able to study a language in the way that a physicist studies material in physics, with very few human prejudices and preconceptions, because the language has to be reduced to its operational characteristics in order to be handled electronically." All this suggests that a new discipline may emerge in which science and linguistics combine to solve international problems of human communication.

It may not be too visionary to suggest that with the perfecting of mechanical translation, significant writings of one land will be made available to interested readers in other countries shortly after they emerge from the presses. Thus, underdeveloped areas of the globe could receive the benefits of advanced science, technology, and knowledge at a rate impossible today. The general problem of the lag in the translation of key works from one language to another will certainly be remedied proportionally as machine translation becomes better developed.